Docking Small Ligands in Flexible Binding Sites

JOANNIS APOSTOLAKIS, ANDREAS PLÜCKTHUN, AMEDEO CAFLISCH

Department of Biochemistry, University of Zürich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland

Received 10 March 1997; accepted 14 July 1997

ABSTRACT: A novel procedure for docking ligands in a flexible binding site is presented. It relies on conjugate gradient minimization, during which nonbonded interactions are gradually switched on. Short Monte Carlo minimization runs are performed on the most promising candidates. Solvation is implicitly taken into account in the evaluation of structures with a continuum model. It is shown that the method is very accurate and can model induced fit in the ligand and the binding site. The docking procedure has been successfully applied to three systems. The first two are the binding of progesterone and 5β-androstane-3,17-dione to the antigen binding fragment of a steroid binding antibody. A comparison of the crystal structures of the free and the two complexed forms reveals that any attempt to model binding must take protein rearrangements into account. Furthermore, the two ligands bind in two different orientations, posing an additional challenge. The third test case is the docking of N^{α}-(2-naphthyl-sulfonyl-glycyl)-D-*para*-amidino-phenyl-alanyl-piperidine (NAPAP) to human α-thrombin. In contrast to steroids, NAPAP is a very flexible ligand, and no information of its conformation in the binding site is used. All docking calculations are started from X-ray conformations of proteins with the uncomplexed binding site. For all three systems the best minima in terms of free energy have a root mean square deviation from the X-ray structure smaller than 1.5 Å for the ligand atoms.
J Comput Chem **19**: 21–37, 1998

Keywords: antisteroid antibody; progesterone; thrombin; NAPAP; flexible docking; MSNI; MCM; finite-difference Poisson–Boltzmann technique

Correspondence to: A. Caflisch; e-mail: caflisch@bioc.unizh.ch
Contract/grant sponsor: Swiss National Science Foundation; contract/grant number 3100-043423.95

Contract/grant sponsor: Swiss Federal Office of Public Health; contract/grant number 3139-043652.95

CCC 0192-8651 / 98 / 010021-17

Introduction

Biochemical specificity relies on the selective binding of molecules to a given protein in a well-defined orientation. It has remained a challenge to predict the structure of ligand–protein complexes theoretically. However, this is the prerequisite for the related but even more difficult task of *de novo* ligand design, that is, the prediction of ligands with high affinity for a given protein target. Many algorithms have relied on a rigid model of the complex, the so-called ''lock and key'' model, with considerable success. The rigid body approach to docking considers only 6 degrees of freedom that can be sampled very thoroughly.[1,2] However, it has been shown that protein and ligand flexibility is important for binding. The induced-fit mechanism[3] proposes that the protein has to undergo usually minor but nevertheless significant conformational changes to accommodate the ligand. The same can be true for the ligand. Allowing flexibility in both the ligand and receptor molecules in the computational model dramatically increases the conformational space that has to be searched.

There are known examples where the empty receptor cannot bind the ligand without a conformational change because the binding site is obstructed.[4] Even in situations where it is possible to dock the ligand to its receptor in a rigid body fashion, it will be important to take induced fit into account in order to optimize the interactions. This is necessary because otherwise the correct structure may not be recognized as such by the scoring function. The interest in flexible docking algorithms emanates also from the expectation that such algorithms can produce reasonable structures, even when starting from low resolution data. This includes not only experimental data but also structures acquired by modeling. The ultimate aim is to model complex structures with little or no experimental information on the particular system. This may be possible by combining homology modeling to obtain the structure of the protein with a flexible docking procedure.

Recently, an incremental construction algorithm based on a tree-search technique[5] and genetic algorithms (GAs)[6,7] have been used for flexible ligand docking. In these docking procedures, the flexibility was limited to the ligand while the binding site was kept fixed. Hence, these calculations are very fast. Jones et al. developed an efficient GA that includes rotational flexibility for the hydrogens and the lone pairs on the hydrogen bonding groups of the receptor.[8]

Zacharias et al. used a Monte Carlo approach with solvation to investigate the conformational changes induced by a mutation in the DNA operator of the λ-repressor–operator complex.[9] During sampling an approximate but efficient model of electrostatics in water was used, while for post-processing of the resulting structures Poisson–Boltzmann calculations were performed. Leach[10] described an algorithm that explores the conformational degrees of freedom of the protein side chains and the ligand. Side chains are allowed to assume different discrete conformations (rotamers) during sampling.[10]

In this article, a new approach is described that in principle can model complete flexibility for the ligand and the receptor. It is based on a combination of minimization with shifted nonbonded interactions (MSNI) and Monte Carlo minimization (MCM). The algorithm seeds the ligand near the putative binding site, allowing structures to occur in which ligand atoms overlap with the protein. In the subsequent minimization, the nonbonded energy terms are modified to avoid high energy gradients. Over the course of the minimization, they are changed back to their original form. This procedure simulates a ligand that gradually feels the field of the protein. Overlap is removed either by a displacement of the ligand away from the protein or, if the ligand has found a pocket where it has low energy, through conformational changes in the protein. The conformational changes that take place during minimization are generally small. For the evaluation of the sampled structures, CHARMM force-field terms[11] are used for the bond energy, the Coulombic interaction *in vacuo*, and the solute van der Waals energy. The finite-difference Poisson–Boltzmann methodology[12–17] is used to calculate the electrostatic free energy of the complexed structures, while nonpolar solvation free energy is approximated by a term proportional to the solvent-accessible surface. The best structures are subjected to the MCM procedure,[18–21] which was expanded to include solvation as described above. The Monte Carlo move randomly perturbs the conformation of the binding site as well as the conformation and orientation of the ligand.

The MSNI/MCM procedure was successfully applied to three systems. The first was the antisteroid antibody DB3.[22] It was shown that this

antibody crossreacts with a number of different steroids and that it binds them in two distinct modes.[4] The two ligands that were chosen for the docking studies, progesterone and 5β-androstane-3,17-dione, are typical for the two binding modes. It is important to note that conformational rearrangement of the antibody binding site is necessary for the binding of the steroids. In particular, the TrpH100 side chain lies on top of the binding site and has to rotate by about 90° to allow access to the steroids. No major domain rearrangement seems to accompany binding.

Since the ligands in this system are rigid by nature, a third test case was chosen where the ligand is completely flexible. The complex between human α-thrombin and N^{α}-(2-naphthyl-sulfonyl-glycyl)-D-*para*-amidino-phenyl-alanyl-piperidine (NAPAP), an archetypal thrombin inhibitor, has been solved to 3.0-Å resolution.[23] The Tyr60A-Pro60B-Pro60C-Trp60D loop covering the S2 pocket of the thrombin binding site assumes slightly different positions in complexes with different inhibitors.[23] It was kept flexible for docking. No information on the conformation of NAPAP was utilized.

Computational Methods

The procedure for docking ligands in a flexible binding site consists of two parts. First the ligand is placed randomly in the binding site, and the ligand–protein complex is minimized with shifted nonbonded interactions to allow gentle resolution of potentially overlapping structures. This is repeated for 1000 seed structures. Then the minima are ranked according to their approximate free energy that includes solvation effects. In the last step, the 20 best minima are subjected to an MCM procedure that includes solvation.

SEEDING

Seeding was performed with the help of CHARMM's built-in random number generator. All random numbers used in this work were taken from uniform distributions. First the initial values of the dihedral angles around rotatable bonds in NAPAP were randomly assigned. Progesterone and 5β-androstane-3,17-dione do not have rotatable bonds. Their structures were taken from the Cambridge Structural Database.[24] Then the ligand molecule was read in, and its center of mass was moved to the center of a 10.0 Å radius sphere that contained the binding site of the protein. In the native DB3 structure (PDB code 1dba), the sphere center was positioned with the graphics package WITNOTP (A. Widmer, unpublished program) between the C_{β} atoms of TrpH100 and TrpH50; in the hirugen–thrombin structure (PDB code 1hgt) the sphere center was positioned between the NH of Gly216 and the S2 pocket. The ligand molecule was then rotated around a randomly oriented vector by a random angle between −180° and 180°. It was then displaced by a randomly oriented vector of random length between 0 and 10 Å. This procedure was repeated to generate 1000 seed structures. To test the dependence of the results on the position of the sphere center, additional MSNI runs were performed for progesterone/DB3 and 5β-androstane-3,17-dione/DB3 with the sphere translated by up to 6 Å. All of these MSNI runs generated two or more structures that were very close to the correct conformation (deviation from the X-ray structure smaller than 1.5 Å) and were among the 20 lowest free energy minima.

MSNI

The seeding leads to structures with considerable overlap between ligand and protein, and in the case of the flexible NAPAP, also between ligand atoms. The nonbonded interactions show an extremely high gradient for overlapping atoms because they grow with the 12th power of the reciprocal distance for the repulsive part of the Lennard–Jones potential. The resulting high gradient can irreversibly damage the structure of the protein and the ligand during minimization. Therefore, the potential is ''shifted'' during the first stages of minimization for the ligand–protein, and in the case of thrombin, intraligand interactions. The square of the interatomic distance (r_{ij}) in the 6–12 Lennard–Jones potential (E_{ij}^{LJ}) is shifted, leading to the following equation:

$$E_{ij}^{LJ} = f\epsilon_{ij}\left[\left(\frac{\sigma_{ij}^2}{r_{ij}^2 + (1-\lambda)\sigma_{ij}^2}\right)^6 - 2\left(\frac{\sigma_{ij}^2}{r_{ij}^2 + (1-\lambda)\sigma_{ij}^2}\right)^3\right]. \quad (1)$$

In the above equation σ_{ij} corresponds to the interatomic distance that yields the minimum value of the unshifted 6–12 potential. The energy at dis-

tance $r_{ij} = \sigma_{ij}$ is $-f\epsilon_{ij}$ for $\lambda = 1$. The parameter λ, which varies between 0 and 1, determines the amount by which the square of the distance is shifted. The additional scaling parameter f, which varies between 0 and 1, is used to further decrease the magnitude of the gradient of E_{ij}^{LJ}. Thus, the softening of the potential is a consequence of the shifting caused by λ and the scaling determined by the factor f. For $\lambda = 1$ and $f = 1$ the potential is equivalent to the CHARMM potential. Similar equations were proposed by Zacharias et al.[25] and independently by Beutler et al.[26] for free energy calculations. Equation (1) allows the shifting and scaling to be controlled independently. A series of test runs showed that the method will also work if only scaling is used. Shifting makes the minimization more stable and thus increases the efficiency of the procedure.

When λ is close to zero the repulsion between interacting atoms is relatively weak and allows considerable overlap. In such cases the Coulombic interactions, even though they are only linearly dependent on the inverse interatomic distance, may also acquire large values and have to be weakened. For this purpose, the following equation is used[26]:

$$E_{ij}^{C} = f\frac{q_i q_j}{4\pi\epsilon\left[r_{ij}^2 + (1-\lambda)^2\right]^{1/2}}. \qquad (2)$$

Note that here the dependence on the shifting term $(1 - \lambda)$ is quadratic. Thus, the shifting of the Coulombic interactions is removed faster than for Lennard–Jones interactions.

The minimization was performed *in vacuo* with a nonbonded interaction cutoff of 5.0 Å and consisted of a loop performing 20 steps of the conjugate gradient algorithm and then an increase of λ by 0.1. To enhance the efficiency of the procedure, it was checked whether the energy still remained favorable after every change in λ. If so, the minimization was skipped and λ increased. Also, if after minimization the Lennard–Jones energy term remained unfavorable, λ was increased only by 0.05 to allow the next 20 iterations to remove the remaining overlap. During all energy minimization procedures, the binding site and the ligand were kept flexible. The definition of the binding site is arbitrary and can in principle include the whole protein. Preliminary runs showed that setting the scaling factor f equal to λ is an efficient way to gradually increase the nonbonded interactions. The last 20 steps at $\lambda = 1$ and $f = 1$ allow a short relaxation with the original CHARMM potential.

It is important to note that no criterion for rejecting structures before minimization is used. Additional bias on the seeding by the initial conformation of the protein is thus avoided. However, it is possible that a ligand is positioned in such a way that it interlocks with the protein. Such starting conformations cannot be resolved by minimization. This problem is not addressed during the MSNI procedure. However, a simple energy evaluation after the MSNI procedure shows whether such a problem has occurred, because strong distortion of bond lengths is characteristic of interlocked structures. High (positive) energy structures are then discarded. A cutoff criterion before MSNI would probably also reject productive starting structures. At the same time, the total computational cost of the algorithm does not increase significantly because MSNI is performed with a low cutoff and is very fast. Although steroids are especially prone to interlocking because of their many rings, it was found that only about 21% and 18% of the starting structures failed to minimize for progesterone and 5β-androstane-3,17-dione, respectively.

The remaining MSNI minima were submitted to a maximum of 500 steps of conjugate gradient minimization with a distance cutoff of 12.0 Å for the nonbonded interactions (i.e., with the full CHARMM22 force field). The minimization was discontinued if the root mean square (RMS) of the energy gradient reached a value of 0.2 kcal/mol Å. The seeding and MSNI procedures were repeated 1000 times to obtain sufficient coverage of conformational space.

ENERGETIC EVALUATION OF MINIMIZED STRUCTURES

The minimized structures were evaluated according to an approximated free energy G, which was calculated as a sum of three terms,

$$G = E_{\mathrm{ff}} + G_{\mathrm{np,\,solvat}} + G_{\mathrm{elect,\,solvat}}. \qquad (3)$$

E_{ff} is the total energy of the CHARMM force field *in vacuo* with no cutoff for the nonbonded interactions. The solvent is treated by a continuum model[13,27–30] and solvation free energy is decomposed into a nonpolar contribution ($G_{\mathrm{np,\,solvat}}$) and an electrostatic contribution ($G_{\mathrm{elect,\,solvat}}$).[31,32] The nonpolar contribution to the free energy is assumed to be proportional to the loss in solvent-

accessible surface area (SAS),[33] on the basis of experimental data on alkane-water partition coefficients.[27] Hence,

$$G_{\mathrm{np,solvat}} = \gamma\left[\mathrm{SAS}^{\mathrm{complex}} - (\mathrm{SAS}^{\mathrm{protein}}_{\mathrm{isolated}} + \mathrm{SAS}^{\mathrm{ligand}}_{\mathrm{isolated}})\right]. \quad (4)$$

The constant γ, which may be interpreted as the vacuum–water microscopic surface tension, is assigned a value of 0.025 kcal/mol Å^2.[34] Only relative values of G are of interest and the absolute scaling does not matter. Therefore, G does not have to be evaluated for the isolated compounds. The solvent-accessible area is computed by an analytical form of the Lee and Richards algorithm[33] as implemented in CHARMM. A probe sphere of 1.4 Å radius was used.

The electrostatic contribution to the solvation free energy ($G_{\mathrm{elect,solvat}}$) is obtained by numerical solution of the linearized Poisson–Boltzmann (LPB) equation. The program used, UHBD,[15,17] is based on a finite-difference scheme that solves the LPB iteratively with the help of a preconditioned conjugate gradient algorithm.[14] The charges are discretized by trilinear weighting.[35] The atomic radii and partial charges were taken from the all-hydrogen CHARMM22 force field. The dielectric interface was defined by the molecular surface, obtained with a solvent probe of 1.4 Å radius. The solute dielectric permittivity was set to 1.0, which is consistent with the CHARMM parametrization of the charges. Also, because the protein is partly and the ligand completely flexible, the dipole reorientation contribution to the dielectric constant is already taken into account to some extent by the minimization. The solvent dielectric was set to 78.5. The dielectric constant was linearly interpolated at the midpoints between grid points intersecting the dielectric boundary (dielectric boundary smoothing) because this reproduces the potential near the discontinuity region more accurately than without smoothing and has been shown to improve convergence.[17,36] Values used were 298 K for the temperature, 100 mM for the ionic strength (corresponding to physiological conditions), and 2.0 Å for the ion exclusion layer.

Standard amino acid protonation states for pH 7.0 were used. This yields a net charge of -1 for Glu and Asp, $+1$ for Lys and Arg, and 0 for the remaining amino acid side chains. His was protonated at N_δ. The N termini were protonated while the C termini were not included in the calculation. At the truncated positions the new termini were treated as carbonyls.

For the LPB calculation a two-step procedure was used. First a grid of $49 \times 42 \times 44$ ($44 \times 44 \times 48$ for thrombin) with a spacing of 2.0 Å was used. This yields a layer of solvent around the protein with a thickness of at least 20.0 Å. The boundary potential was obtained by treating the complex as a single Debye–Hückel sphere with the molecule's net charge. A second, focused calculation was performed on a $77 \times 64 \times 68$ ($67 \times 67 \times 76$ for thrombin) grid with a grid spacing of 1.0 Å, which yielded a solvent layer around the protein with a thickness of at least 10.0 Å. The boundary potential for the fine grid was obtained by interpolation from the coarse grid.

$G_{\mathrm{elect,solvat}}$ was calculated as follows. Both the finite-difference approximation to the Coulombic interaction energy between charged atoms *in vacuo* and the interaction energy of each atom with its own potential (this contribution arises from the discretization of the atomic charges on a grid) were subtracted from the total electrostatic energy of the system calculated by the finite-difference LPB technique.[37] These three energy terms are calculated on the same grid to obtain consistent results. When the solute dielectric is 1.0, this yields the same result as the usual (and computationally more expensive) method of performing two finite-difference calculations: the first with the dielectric solute in a high dielectric continuum and the second with the low dielectric solute *in vacuo* (solvent dielectric of unity and zero ionic strength). The position of all grid points is kept fixed relative to the rigid part of the protein. It should be noted that this includes the fixed atoms contribution to the self-energy that is the major source of grid error.

METROPOLIS MCM WITH SOLVATION

Among the 1000 MSNI minima the 20 best in terms of free energy were subjected to 200 cycles of MCM. The 10 MCM runs that yielded the lowest energy structures were continued for another 200 MCM cycles. The MCM procedure allows the protein to relax with the ligand in the binding site to obtain the conformational changes that the ligand induces in the protein and, at the same time, decreases the energy of the complex.

The MCM scheme was originally applied to find the global minimum energy conformation of oligopeptides.[18,19] The MCM procedure used in this work is based on a previously developed approach for docking[20,38] that consisted of random perturbations of the ligand torsional degrees of

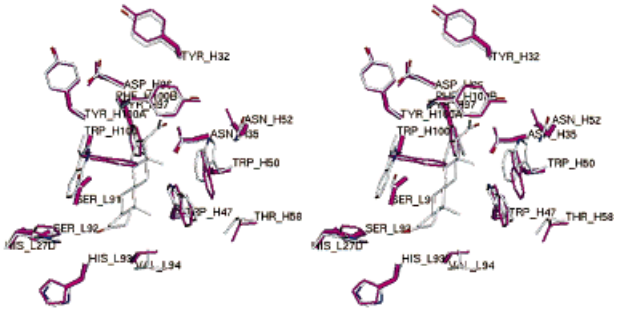

FIGURE 1. Stereoview of the progesterone / DB3 binding site (colored by atom type, PDB code 1dbb) and the native DB3 structure (magenta, PDB code 1dba). For the sake of clarity, only the residues that are perturbed by an MC move are shown. The bonds whose torsion angles are affected by an MC move are shown in thick lines in the native structure.

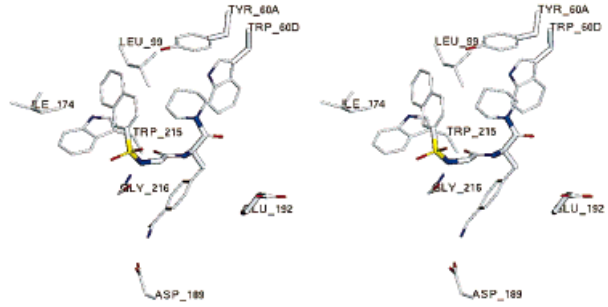

FIGURE 2. Stereoview of the NAPAP / thrombin complex (PDB code 1dwd). The bonds whose torsion angles are affected by an MC move are shown in thick lines.

freedom and subsequent conjugate gradient minimization during which the protein was kept rigid. The Metropolis criterion was applied to the energy of the resulting structure. In a more recent docking study[21] and in this work, besides the orientation and overall position of the ligand, the torsional degrees of freedom of the protein side chains are also perturbed in each move. Among all of the rotatable bonds in the ligand and in the binding site side chains (Figs. 1, 2), a number n were randomly selected and perturbed in each move, where n is a variable chosen with a probability 2^{-n}.[19] In addition, the ligand was rotated as a rigid body by an angle between $-30°$ and $30°$ about a randomly oriented axis going through its center of geometry. It was also displaced by a randomly oriented vector of random length between 0 and 0.35 Å.

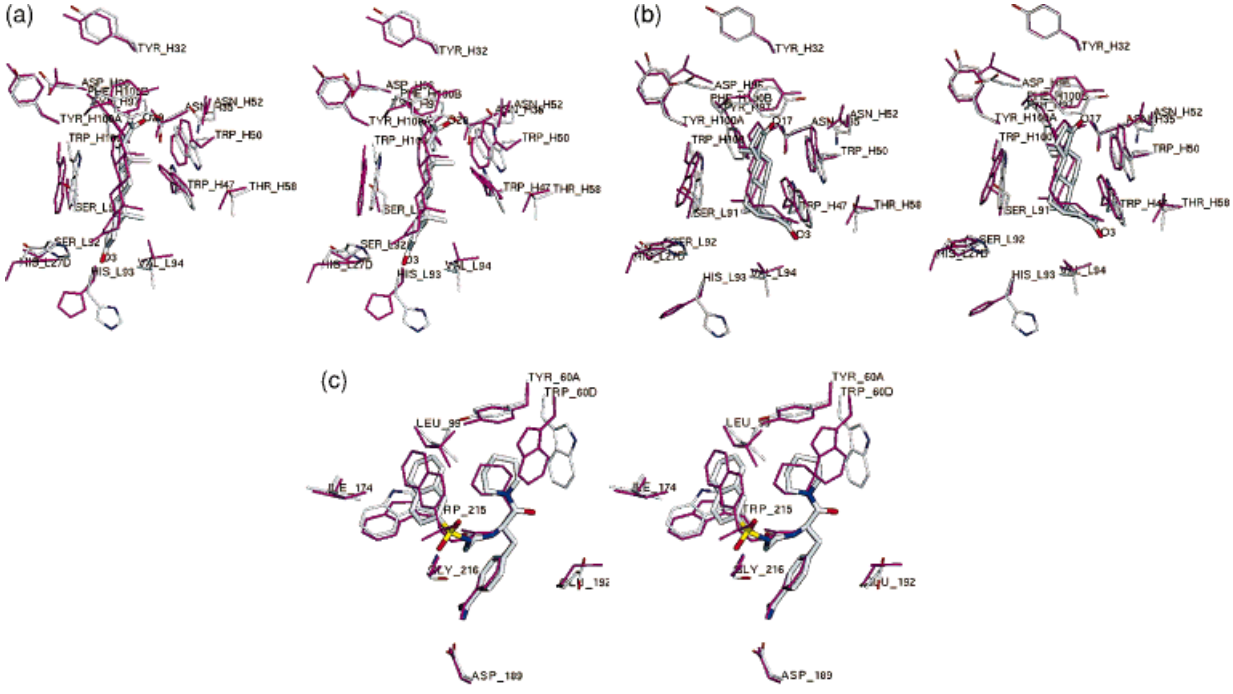

FIGURE 6. MSNI / MCM minima with the lowest binding free energy (colored by atom type) and X-ray structure (magenta). (a) Progesterone / DB3 complex, (b) 5β-androstane-3,17-dione / DB3 complex, and (c) NAPAP / thrombin complex.

After each move, 10 steps of steepest descent minimization were performed to remove steric clashes. Structures with force-field energy larger than 0 kcal/mol were discarded at this point. Conjugate gradient minimization was performed until the RMS of the energy gradient was less than 0.2 kcal/mol Å or a maximum of 200 steps was exceeded. The main-chain and side-chain atoms in the binding site of the protein and the atoms of the ligand were flexible during the minimization to enhance sampling. The CHARMM22 default cutoff value for the nonbonded interactions was used (12.0 Å), and the constant dielectric was set equal to 1.

The minimized structures were accepted, depending on their approximate free energy [eq. (3)], with a probability $p = \min(1, e^{-\Delta G/kT})$.[39] The major improvement of the MCM procedure used in this work is that the Metropolis criterion is applied to the approximate free energy in solution and not to the force-field energy alone. Hence, the sampling is directed toward conformations with optimal free energy rather than optimal vacuum energy as calculated by CHARMM.

For kT a value of 1.0 kcal/mol was used, which corresponds roughly to 500 K. No attempt was made to optimize the method with respect to the temperature of the MCM run. At the end of the 10 400-cycle MCM runs all of the minima within 10 kcal/mol of the best free energy minimum were clustered by their ligand RMS deviation (RMSD) using the program GATHER[40] with a 0.75 Å cutoff. GATHER sorts the minima according to energy and selects the lowest energy minimum as representative of the first cluster. Minima that differ from the representative by less than the cutoff are included in the cluster. The procedure is repeated by selecting the next minimum, which is not already a member of a cluster, until all ligands have been included in a cluster.

TEST PROBLEMS

Antisteroid Antibody DB3

The monoclonal antibody DB3 binds progesterone with nanomolar affinity ($K_D = 1$ nM).[4] At the same time, it crossreacts with similar affinity with a number of structurally distinct steroids. The structures of the antigen binding fragment of DB3 in free form and in complex with five different steroid ligands have been determined.[4, 22] There are two striking features that are important for this work. The first relates to the rotation that the side chain of TrpH100 undergoes upon binding (Fig. 1). Comparison of the native conformation and five complexes with different steroids shows that the binding site is occluded by the indole ring of TrpH100 in the unliganded structure. In keeping with the terminology used by Arevalo et al.,[4] this conformation is referred to as the closed form of the binding site; the bound conformation where the ring gives way to the ligand and forms part of the wall of the binding pocket is called the open form. This induced fit makes it impossible to predict the structure of the complex using a rigid protein in the unbound conformation. Neither would this be possible with a limited flexibility of the binding site, for example, allowing just rotations of the hydrogen bond donor and acceptor groups in the side chains as in ref. 8. To test the present docking procedure, the ligand-docking calculations were started from the closed form of the DB3 antibody and the results compared with the X-ray structure of the complex (Table I).

The second feature is that there are two distinct binding modes for different ligands. The two modes differ in the orientation of the ligand in the binding pocket. The two test cases considered in this work, progesterone and 5β-androstane-3,17-dione (Fig. 3a, b), exemplify this. Progesterone

TABLE I.
X-Ray Structures Used for Docking and Calculation of RMSD.

System	PDB Code	Resolution	Reference
DB3 native	1dba	2.8	22
Progesterone / DB3	1dbb	2.7	22
5β-Androstane-3,17-dione / DB3	1dbk	3.0	4
Hirugen / human α-thrombin	1hgt	2.2	58
NAPAP and hirudin / human α-thrombin	1dwd	3.0	23

The resolution is given in Å. The structures used for docking are 1dba for the DB3 antibody and 1hgt for human α-thrombin. The structures 1dbb, 1dbk, and 1dwd were used to calculate the RMSD.

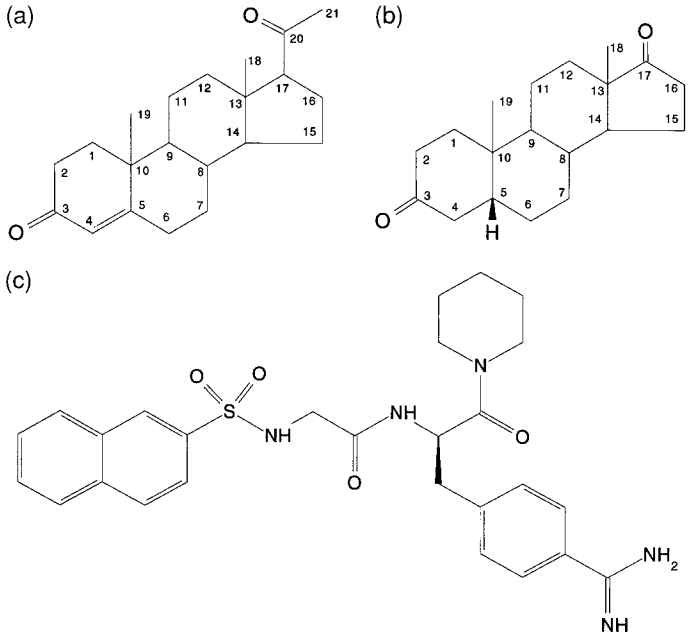

FIGURE 3. Structural formulas of the ligands. (a) Progesterone (b) 5β-androstane-3,17-dione, and (c) NAPAP.

binds with the steroid β face pointing toward TrpH50 whereas 5β-androstane-3,17-dione binds with the steroid β face in the opposite direction.[4] It should be noted here that the two ligands are significantly different in terms of 3-dimensional structure. Progesterone is relatively flat, while 5β-androstane-3,17-dione has a stereocenter at atom C5 that results in the A ring being almost perpendicular to the rest of the molecule. The two possible binding modes present an interesting test case for the quality of the energy function used to evaluate the minima. One expects that the docking algorithm will find both orientations. Thus, it is interesting to test if the energy function is able to properly rank the two binding modes for the two ligands.

During the minimization, the binding site atoms are kept flexible while the rest of the protein is fixed. The definition of the binding site is somewhat arbitrary; to introduce as little bias as possible toward the complex structure, the following approach was taken. The unliganded and the complex structures were superimposed and then a selection was performed to include all main-chain and side-chain atoms in the unliganded structure that were within 8 Å of any heavy atom of progesterone. This yielded 872 flexible atoms in the protein. For the docking of 5β-androstane-3,17-dione, the same atoms were kept flexible. Visual analysis with the program WITNOTP confirmed that these atoms form the binding site of the antibody DB3. For the calculation of the energy only the Fv part of the antibody was included.

Human α-Thrombin

Thrombin is a trypsinlike serine protease that plays an important role in hemostasis and thrombosis.[41] Its crystal structure in complexes with a series of inhibitors is known.[42] However, the structure of the unbound protein is not available. For the docking studies, the thrombin–hirugen complex was used as the starting structure (Table I). Hirugen binds at the so-called exosite and not at the catalytic center. Thus, there is no positive bias in the structure for the binding of the active site inhibitor NAPAP (Fig. 3c). The benzamidine moiety of NAPAP occupies the S1 pocket; its piperidine ring fits in the S2 pocket of the active site, and the naphthyl group fills the S3 pocket. In addition, the glycyl group is involved in two hydrogen bonds with the main-chain NH and CO of Gly216 (Fig. 2).

The conformational changes of thrombin upon binding of NAPAP are relatively small, with residues in the active site moving 0.5–1.0 Å. Nevertheless, the loop Tyr60A-Pro60B-Pro60C-Trp60D in S2 assumes different orientations, depending on the size of the group in S2.[23] The binding site was

defined in the way described for the DB3 antibody and consisted of 1189 atoms.

COMPUTATION TIMES

On average, one MSNI cycle requires 10 min while an MCM cycle requires 6 min of CPU time on an SGI workstation with a R4400/250 MHz processor. On the same processor the UHBD calculation takes about 3 min of CPU time. The complete docking required about 5 days of CPU time on an eight R4400/200 MHz processor SGI challenge.

Results

The 20 best MSNI minima were subjected to the MCM procedure. After MCM and GATHER, there were six, seven, and five different minima within 10 kcal/mol of the best free energy structure for progesterone/DB3, 5β-androstane-3,17-dione/DB3, and NAPAP/thrombin, respectively. These structures are discussed with particular emphasis on the energy contributions and solvation free energy values. For the calculation of the RMS deviation between the minima and the X-ray structures, the conformations were fitted by superimposing the rigid part of the protein. Unless otherwise stated, the values given for the RMSD refer to the values for the ligand because these are the most representative for the quality of the structure. Including the flexible atoms of the protein in the RMSD generally leads to low values with low variance, because the flexible atoms of the protein that are not randomly perturbed are not displaced significantly.

MSNI

Progesterone/DB3

The RMSD between the docked conformations of progesterone and its conformation in the crystal structure is very low for 3 of the 10 structures with the best free energies after MSNI. There are three structures in which progesterone is flipped by 180° around an axis perpendicular to its longest dimension with respect to the crystal structure of the complex. This fact is due to the approximate symmetry of progesterone. The minima with the sixth, eighth, and ninth lowest free energy after MSNI exhibit this ''inverse'' binding, which is different from both the progesterone and 5β-androstane-3,17-dione/DB3 binding modes.

5β-Androstane-3,17-Dione/DB3

Most of the low energy structures have 5β-androstane-3,17-dione bound in an orientation that is rotated with respect to the crystal structure by about 120° around an axis along the longest dimension. In this orientation, which is different from the correct binding modes of both progesterone and 5β-androstane-3,17-dione, the two methyl groups of the steroid molecule are exposed to solvent. The 12th minimum is close to the binding mode of progesterone.

NAPAP/Thrombin

Thrombin is known not to rearrange dramatically upon binding of different inhibitors.[23] In this study, we were interested in studying the effect of flexibility on the ligand molecule. When the ligand is small, the algorithm will allow a thorough sampling of the conformational space that is available to it. All of the NAPAP rotatable bonds (Fig. 2) were randomized. Therefore, completely random conformations and orientations in the binding site were produced for the NAPAP inhibitor. This procedure simulates a ligand that has no preferred (or an unknown) conformation in solution and becomes constrained upon binding to a protein. One would also need to take this approach when induced conformational changes in the ligand are large.

It was found by visual inspection that the piperidine ring was placed in the S2 pocket in 8 and the benzamidine occupies the S1-pocket in 17 of the best 20 MSNI minima. Yet, the β-naphthyl moiety of NAPAP did not occupy the S3 pocket; that is, it was not placed correctly in the 20 best MSNI minima.

Figure 4 shows a plot of the total free and the force-field energy relative to the RMSD of the ligand from the X-ray structure of the complex. The overall correlation is relatively low because the energy hypersurface is rugged with many different local minima. The proximity of a local minimum to the global one correlates only to some extent to its energy. The correlation coefficient is similar for the free energy versus RMSD (0.38) and the force-field energy versus RMSD (0.36) if all structures are taken into account. However, the correlation coefficient for the 100 best structures

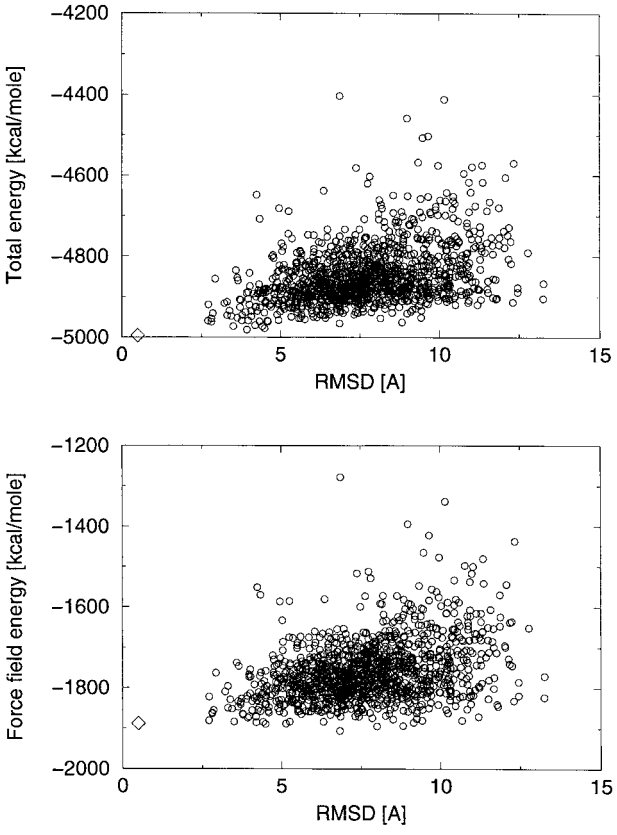

FIGURE 4. Scatter plot of the energy (kcal / mol) relative to the RMSD (Å) from the crystal structure of the NAPAP / thrombin complex. The data shown are for the structures obtained by MSNI. NAPAP was also docked manually in the ''free'' conformation of human α-thrombin so as to model the conformation found in the crystallographic structure of the NAPAP / thrombin complex. The energies of the resulting structure are denoted by diamonds. It is not possible to use energy values calculated from the X-ray structure of the NAPAP / thrombin complex because of the different conformation of the protein part not involved in the binding.

according to free and force-field energy is 0.42 and 0.12, respectively. The plot does not indicate a significant advantage for the free energy function over the force-field energy. However, it is necessary to include solvation when protein flexibility is allowed in the docking calculations to avoid sampling irrelevant regions of the conformational space. Figure 4 shows that MSNI generates highly favorable conformations over a wide range of RMSD. This is important for the sampling of alternative binding modes (see below).

MCM

The plot of the energy versus the number of accepted structures is shown in Figure 5 for the MCM run that yielded the lowest free energy minima of NAPAP.

The shape of the plotted curves is typical, and most of the MCM runs have similar behavior. The electrostatic solvation energy shows relatively large fluctuations whereas the nonpolar solvation contribution is almost constant. Both the force-field energy and the total free energy, which is used in the Metropolis criterion, improve significantly, although new structures with less favorable total free energy are sometimes accepted due to the kT value of 1.0 kcal/mol (see Computational Methods section). The RMSD for the ligand atoms of the lowest energy structure to the initial structure of the run lies usually between 2 and 3 Å. The relatively low displacement from the starting structures indicates that the conformational search is much more efficient when performing short runs with different initial conditions (i.e., different initial conformations of the complex).

Antisteroid Antibody DB3

Although the MCM was performed at 500 K, the acceptance ratio was relatively low at about 20%. This relatively low acceptance ratio might be a consequence of the random perturbations used in the present implementation, which are not expected to retain favorable interactions in a highly cooperative system such as a protein. Another reason for the low acceptance may be the fact that solvation is taken into account in the Metropolis criterion. It is not clear how to design a move that efficiently leads to structures with favorable solvation energy. On the other hand, there is a dramatic improvement of the energy in MCM. On average, it improves by 42 and 39 kcal/mol for the 10 progesterone and androstane runs of 400 MCM cycles.

The combination of MSNI and MCM shows convergence for progesterone and androstane (Table II). It is interesting to note that an MSNI minimum of progesterone with an RMSD from the X-ray structure of 2.9 Å converged to a structure with an RMSD from the X-ray structure of only 1.0 Å after MCM (minimum 4 in Table II). MCM was determinant for the convergence of 5β-androstane-3,17-dione.

One striking feature that is common among all of the structures found for the antigen binding

fragment of the DB3 antibody and distinguishes them from the crystal structures is that the loop L1 moves toward the binding site by approximately 1 Å. The structure is thus somewhat more compact. We consider this an artifact of the minimization being performed *in vacuo*. This conformational change does not directly affect the orientation of the steroid molecule in the binding site.

Progesterone/DB3. Figure 6a shows the lowest free energy minimum obtained with the MSNI/MCM procedure and the X-ray conformation. The agreement is excellent; this is also evident in the ligand and total RMSD values of 0.9 and 1.5 Å, respectively. According to the crystallographic analysis, the two residues that move the most upon binding are TrpH100 and AsnH99. In all of the progesterone/DB3 minima obtained by MCM, the AsnH99 side chain is placed somewhat closer to its position in the complex conformation than in the native structure (not shown). However, the amide group of the side chain is rotated by 180° with respect to the crystal structure. This discrepancy is not important because AsnH99 points away from the binding site and is solvated. Furthermore, the assignment of amide groups is usually made based on energetic considerations at these resolutions because the carbonyl and the amine group show similar electron densities. In all minima, the TrpH100 side chain is displaced into the correct conformation: it undergoes a rotation of about 90° around χ_2 to open the binding site.

An interesting point is that in contrast to the crystal structure none of the minima exhibits a hydrogen bond between the carbonyl oxygen at C3 and the side chain of HisL27d. This is due to the fact that the His was protonated on N_δ in the docking calculations, because in the uncomplexed conformation of the antibody this is more favorable. With such protonation the HisL27d N_δ donates to the carbonyl of SerL92, whereas the X-ray structure of the complex suggests that the N_ϵ of HisL27d is protonated. In the minimized structures, the proton on the N_δ of HisL27d forms a hydrogen bond to the main-chain carbonyl of

TABLE II.
Lowest Free Energy Minima.

		Solvation Energy			MCM		MSNI	
	E_{ff}	G_{elect}	G_{np}	G	Ligand RMSD	Total RMSD	Ligand RMSD	Total RMSD
Progesterone / DB3 complex								
1	**−1304.2**	**−1808.1**	**193.6**	**−2918.7**	0.9	1.5	1.4	1.6
2	5.1	3.5	0.0	8.6	1.2	1.4	1.2	1.2
3	14.2	−4.4	−0.9	9.0	0.9	1.2	1.7	1.1
4	9.1	0.9	−1.0	9.0	1.0	1.2	2.9	1.1
5	4.6	4.8	−0.0	9.3	1.3	1.5	1.2	1.2
5 β-Androstane-3,17-dione / DB3 complex								
1	**−1249.2**	**−1810.4**	**193.5**	**−2866.2**	0.9	1.3	3.6	1.6
2	2.5	−1.1	0.2	1.6	0.7	1.3	3.6	1.6
3	4.6	−1.5	−0.1	2.9	1.1	1.3	3.6	1.6
4	1.5	2.1	−0.0	3.6	0.5	1.3	3.6	1.6
5	10.2	−6.2	−0.2	3.8	3.5	1.7	3.6	1.6
NAPAP / thrombin complex								
1	**−1888.8**	**−3331.1**	**195.2**	**−5024.7**	1.4	1.5	2.8	2.1
2	1.8	1.7	−0.1	3.4	5.2	3.9	4.2	3.1
3	6.2	−3.2	0.9	3.9	4.1	3.1	4.0	2.9
4	−0.2	5.3	−0.7	4.4	5.0	3.8	4.2	3.1
5	−4.8	9.4	0.1	4.7	5.5	4.1	4.2	3.1

E_{ff}: force-field energy; G_{elect}: electrostatic solvation energy (PB); G_{np}: nonpolar solvation energy (SAS); G: free energy.
Energy values are in kcal / mol and RMS deviations in Å. The absolute energy of the lowest minima is shown in bold. The energy terms of all other minima are given relative to those of the lowest energy minimum. RMSD total gives the RMSD between calculated and X-ray structure for the atoms in the ligand and those in the protein that are perturbed in the Monte Carlo move. Comparing all flexible atoms for the RMSD calculation leads to low values for the RMSD and low variance because the flexible residues further away from the ligand do not move much and the RMSD is an average quantity. Columns six and seven contain RMSD values after MCM (i.e., at the end of docking), while the last two columns contain RMSD values after MSNI.

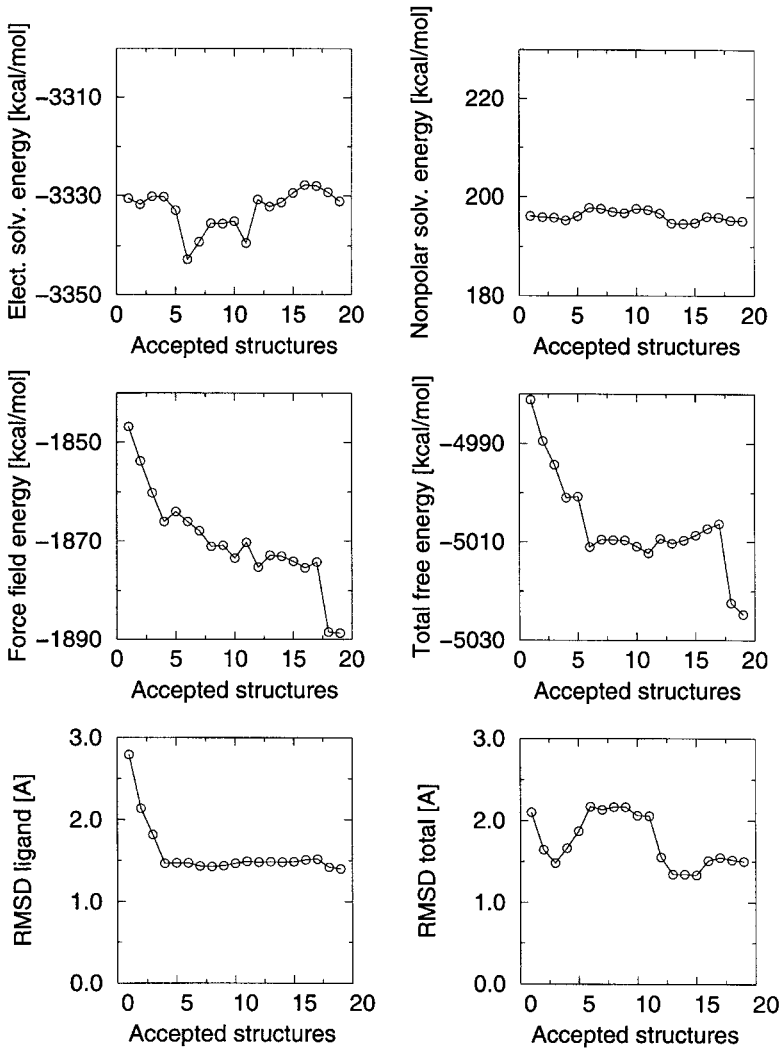

FIGURE 5. Plots of energy (kcal / mol) and RMSD (Å) for the MCM run that yielded the lowest energy minimum of the NAPAP / thrombin complex. Upper row left: electrostatic solvation energy ($G_{\text{elect, solvat}}$). Upper row right: nonpolar solvation energy ($G_{\text{np, solvat}}$). Middle row left: force-field energy (E_{ff}). Middle row right: total free energy (G). Lower row left: root mean square deviation between the sampled structures and the X-ray structure for the ligand. Lower row right: root mean square deviation of the ligand and the side chains that are perturbed by an MC move. All plots are over the number of accepted structures.

SerL92 or the imidazole nitrogen of HisL93. This may indicate that the formation of this hydrogen bond is not necessary for the orientation of the progesterone molecule in the binding site. In addition, because this hydrogen bond is exposed to solvent it is not expected to be very important for binding affinity. In the lowest energy minimum (number 1 in Table II), the distance between the C_ϵ of HisL27d and the O3 of the progesterone is 3.0 Å (Fig. 6a), equal to the distance between the N_ϵ and O3 in the crystal structure. The TrpH50 side chain lies in the same plane as in the crystal structure but in a different orientation (Fig. 6a). It would constitute an overinterpretation of the crystallographic data as well as ours to discuss the significance of this discrepancy. In every other respect, we consider the differences between the crystal structure and the first minimum to be within experimental error. All of the five best minima have a ligand RMSD smaller than 1.3 Å (Table II).

5β-Androstane-3,17-Dione/DB3. The 5β-androstane-3,17-dione structure with the lowest free energy is very similar to the crystal structure in terms of RMSD (Table II) and of interactions with the protein. Figure 6b shows that the lowest en-

ergy minimum overlaps the X-ray structure of the complex. The success of this docking experiment was somewhat unexpected. The reason for this is that in the crystal structure of the complex there is a water molecule mediating an indirect interaction between the carbonyl O3 of the steroid and ThrH58.[4] A possible explanation is that the PB calculation implicitly places a water at the relevant position by filling the empty space with a high dielectric medium.

Minima numbers 1–4 are close to the X-ray structure (ligand RMSD smaller than 1.1 Å, Table II). Minimum 5 binds in the same orientation as progesterone. Because of the N_δ protonation of HisL27d (see previous subsection), 5β-androstane-3,17-dione cannot form a hydrogen bond to HisL27d in minimum 5, but again there is a van der Waals contact between the carbonyl O3 on the steroid molecule and the C_ϵ of HisL27d. It is important to note that for both progesterone and 5β-androstane-3,17-dione the docking procedure finds the three relevant binding modes: the two binding modes of the crystal structures and the inverse orientation. Moreover, the energy function yields the correct ranking.

NAPAP / Thrombin

In all of the five best minima, the benzamidine is placed in S1. The conformation with the lowest free energy is the closest to the crystal structure of the complex (Table II). However, there are a few notable differences as can be seen in Figure 6c. The naphthyl moiety is rotated around the bond to the sulfur by 180° with respect to the X-ray structure. This does not result in any change in the interactions with the protein. It is interesting to note that in the 3.0 Å resolution crystal structure of the complex it was not possible to distinguish which of the two alternative conformations of the naphthyl is preferred.[23] A more recent study of the complex between NAPAP and bovine α-thrombin shows that the naphthyl moiety has an orientation identical to the one of the X-ray structure shown in Figure 6c.[43] For the calculation of the RMSD, the structure with the PDB code 1dwd[23] was used, which is also the structure shown in Figure 6c. The piperidine moiety is also rotated by 180° around the bond between the N and the carbonyl C. This may be the reason why Trp60D is not lying on top of NAPAP but is more solvated. The most significant discrepancy with respect to the crystal structure is the exposed NH of the sulfonamide, which does not form an H-bond with the CO of Gly216 as in the crystal structure.

In minima 2, 4, and 5, NAPAP is docked in an orientation very close to the *second inhibitor binding mode* first discovered in an X-ray structure of a D-phenyl derivative of NAPAP (compound **3c** in refer. 44). The naphthyl group is positioned between the S2 and S3 pocket, while the piperidine is in a rather exposed position at the front of the S3 pocket. The main difference between the second inhibitor binding mode and the orientation of NAPAP in minima 2, 4, and 5 is the lack of the hydrogen bond between the sulfonamide oxygen and the NH of Gly216 in the NAPAP minima.

Minimum 3 is very close to minimum 1, apart from the orientation of the naphthyl group that points toward the solvent in minimum 3. The different orientations of the naphthyl group in minima 1, 2, and 3 suggest that the affinity of the naphthyl group for the S3 pocket is not very high and that the intraligand hydrophobic interactions between the naphthyl and piperidine moieties are determinant for the high affinity of NAPAP for thrombin in agreement with experimental data.[45]

The results show that the structure with the best free energy lies within 1.5 Å of the complex conformation for every test case. Furthermore, the calculated approximate free energy was able to identify the best minima correctly, despite the large degree of flexibility that leads to a high-dimensional smooth energy hypersurface. This vastly increases the number of local minima and makes the identification of the global one much more difficult.

Discussion

SIGNIFICANCE OF SECONDARY BINDING MODES

The importance of secondary binding modes may not be directly evident. For example, a binding mode with a free energy of binding of 1 kcal/mol higher than the global minimum is less populated by a factor of roughly 5 and therefore is much more difficult to observe experimentally. However, secondary binding modes may be important for enzyme action because they might represent intermediate bound states on the way of the substrate to the reaction center. In addition, in drug design the example of thrombin inhibitors shows that a binding mode that does not correspond to the one preferred by the substrate may turn out to be better suited for inhibitor binding[44]

and, more importantly, there might be completely different binding modes for closely related ligands.[44,46] Another example where secondary binding modes play a role is in the addition of groups that destroy the interactions of a ligand with a protein to increase cross-reacting selectivity. This may fail if the ligand has a secondary binding mode whose binding is not interfered with by the additional group.

In the case of the antisteroid antibody, which is known to bind different steroids in distinct (opposite) orientations, it is possible that the same ligand can also be bound in different ways. Even though alternative binding modes influence the affinity only weakly in most cases, they may occur in this situation because of the small size and the approximate symmetry of the steroids. As was shown in the Results section for the three test systems, the MSNI/MCM docking procedure sampled the relevant minimum energy conformations.

COMPARISON WITH CURRENTLY AVAILABLE METHODS

A number of docking procedures were recently described[5–7,47–49] that can handle flexibility in the ligand. They use either a grid approach[49] or a tree-search technique[5] or GAs. Because the evaluation of the energy is the speed limiting factor for docking algorithms, programs like DOCK[2] and AutoDock[49] have been developed that precalculate grids of interaction energy based on the macromolecular target. The interaction energy for a specific conformation is then calculated by interpolation between neighboring grid points. AutoDock then performs simulated annealing in the space of the flexible ligand. These methods have been shown to be efficient and reliable. However, the grid-based approach assumes purely additive interactions for the ligand. This is not valid for the solvent screening of the electrostatic interactions. It is also not yet quite clear how to introduce protein flexibility in grid-based methods.

GAs offer a very efficient possibility for tackling general optimization problems when a complete search lies outside the scope of the computational capacity. Possible solutions of the problem are encoded in data structures called chromosomes.[50] The chromosomes are optimized through genetic operators, like mutation and crossover, and the fittest are selected according to a scoring function to propagate into the next generation. It is important to compare the MSNI/MCM approach with a recently developed docking procedure that includes some limited flexibility in the binding site, namely in all torsion angles to hydrogen bond donors or acceptors.[8] The chromosomes used in the GA of Jones et al. encode the orientation of the ligand by way of the hydrogen bonds that it can form with the protein.[8] The authors suggest that in principle any degree of flexibility can be obtained through such coding. This might, however, be problematic because randomly changing backbone or side-chain torsional degrees of freedom leads to overlapping structures with extremely poor energy. Moreover, Jones and coworkers do not include the intraprotein Lennard–Jones term in the fitness function.[8]

Jiang and Kim have proposed a method for ''soft'' docking that implicitly allows for induced fit or changes in the conformation of receptor and ligand.[51] The method describes molecules with the help of low resolution surface cubes that are used to find relative orientations of the molecules that exhibit high complementarity. As remarked by the authors, it might be possible to combine such approximate methods with more detailed ones to increase efficiency on the one hand and resolution on the other.

Exploring the side-chain flexibility of the receptor protein with the help of rotamers is another efficient method.[10] Its obvious limitation of the fixed backbone may in many cases not be very important. However, the flexibility of the side chains is further limited by the rigidity of the backbone in a closely packed structure such as a protein.

Jackson and Sternberg recently brought attention to the problem of the discrimination between near-native and nonnative structures.[52] They report that most algorithms will produce nonnative structures that have similar or even better energies than the near-native ones. They attribute this problem solely to the quality of the evaluation function. There is no doubt that the evaluation function used is of supreme importance. However, our results indicate that there may be an additional factor that needs to be taken into account, namely the completeness of relaxation. This is best seen by the comparison between the results of MSNI alone and MSNI/MCM. By allowing the protein to relax with the ligand in the binding site during MCM, the energy decreases dramatically. Because a significant correlation between proximity to the global minimum and energy is expected, significant relaxation of the macromolecular assembly may be

necessary to discriminate near-native from nonnative structures. From our results, conventional energy minimization does not appear to be sufficient.

An interesting implication of the results in the NAPAP/thrombin docking study is that some of the groups (in this case the benzamidine and the piperidine) in a flexible ligand seem to find the correct position independently. A number of ligand design algorithms are based on the assumption that all the groups of a ligand will do so; fragments of the ligand are first docked individually and then connected to yield a complete molecule.[40,53–57] On the other hand, such algorithms do not take the intraligand interactions into account. In the case of NAPAP, for example, the hydrophobic interaction between the naphthyl and piperidine cannot be neglected and seems decisive in determining the bound conformation of NAPAP.

The MSNI/MCM procedure showed convergence for the progesterone/DB3 and the 5β-androstane-3,17-dione/DB3 complex, even though only 1000 conformations were used as starting structures. It was suggested by an anonymous reviewer that the efficiency in sampling might rely on the softening of the overlap potential in MSNI. This allows every starting structure to inherently sample more conformational space.

The main advantage of the MSNI/MCM method is that it allows unrestricted flexibility; and by using the ligand to effect the induced changes in the protein, a correlated sampling in the degrees of freedom of the problem is achieved. That means that only the part of the protein that is close to the ligand in every structure changes its conformation, accommodating the ligand. The method automatically limits the conformational flexibility in the vicinity of the docked ligand. The MCM procedure then allows conformational strain to dissipate and be partly eliminated at the surface or other flexible parts of the protein. Correlations in sampling exist also in GA methods where they are effected by genome structure. This is, however, an additional parameter that makes the methods dependent on subjective criteria.

The second point that distinguishes this work from most other docking studies is the use of a free energy with a more rigorous determination of the solvation free energy instead of a faster but less accurate scoring function or *in vacuo* energy. The use of a more elaborate energy function is necessitated by the inclusion of protein flexibility. Methods used to dock flexible ligands in a rigid protein do not need to calculate the intraprotein energy. Furthermore, most of these methods are successful only if they can use the bound conformation of the protein. In the conformation of the complex geometric complementarity is maximal, and thus a simple scoring function that emphasizes geometric features will usually find the correct minimum. In such situations, Lennard–Jones interactions are heavily dependent on perfect fit and will obscure the effect of electrostatics and solvation. With a flexible protein, however, the dimensionality of the problem is increased. This results in a smoother energetic landscape where the contribution of other energetic terms becomes more important. To conclude the argument, it can be said that the more flexibility is allowed, the more rigorous and detailed the energy function has to be.

It is important to note that the intrasolute entropy contribution to the free energy is completely neglected in the present work. This term is often omitted in docking studies, probably because its rigorous calculation would involve the creation of ensembles of conformations around the proposed structures. Such extensive sampling is computationally prohibitive for a large number of structures. The neglect of the intrasolute entropy is partly justified in the case of the antisteroid antibody because of the rigidity of the ligands. That does not mean that entropic factors related to protein flexibility may not be important for the affinity and the structure of the complex. It has been suggested that loss of side-chain conformational entropy is a second-order effect.[52]

The proposed method is relatively easy to apply and can be generalized for the use with any forcefield and PB solver. The combination of different approaches to docking (in this work MSNI and MCM) seems to be advantageous because different algorithms can compensate for each others weaknesses.

The MSNI/MCM docking method yielded the correct binding mode for the progesterone/DB3 antibody complex, the 5β-androstane-3,17-dione/DB3 complex, and the NAPAP/thrombin complex starting from the native conformation of the protein and, in the case of NAPAP, without foreknowledge of the structure of the ligand.

Substrate binding to the enzyme binding site might be investigated with the MSNI/MCM docking procedure. By supplementing the description of the molecular system with degrees of freedom corresponding to a reaction coordinate, it should

be possible to model transition state binding. In the context of the transition state preferential binding model, it would be interesting to establish a method to determine transition state complexes based only on energetic criteria.

Acknowledgments

This work was supported by the Swiss National Science Foundation (Schweizerischer Nationalfonds, Grant 3100-043423.95) and the Swiss Federal Office of Public Health (Nationales Aids-Forschungs-Programm, Grant 3139-043652.95). We thank Dr. C. Ehrhardt for helpful discussions and interesting comments on the manuscript and R. Jorissen and Dr. A. Mark for interesting discussions. The release 4.1 of the UHBD program, used for all finite-difference Poisson–Boltzmann calculations was kindly provided by Prof. J. A. McCammon. Visual analysis was performed with the molecular modeling program WITNOTP developed and kindly provided by Dr. A. Widmer, Novartis Pharma Inc., Basel. WITNOTP was also used to prepare Figures 1, 2, and 6a–c. The calculations were performed on an SGI Indigo2 (R4400) and an SGI Challenge (with eight R4400 processors). The UNIX, CHARMM, and UHBD scripts are available from the first author (e-mail: joannis@bioc.unizh.ch).

Nomenclature

CHARMM	Chemistry at Harvard Macromolecular Mechanics
CPU	central processing unit
DNA	deoxyribonucleic acid
GA	genetic algorithm
LPB	linearized Poisson–Boltzmann
MCM	Monte Carlo minimization
MSNI	minimization with shifted nonbonded interactions
NAPAP	N^{α}-(2-naphthyl-sulfonyl-glycyl)-D-*para*-amidino-phenyl-alanyl-piperidine
PB	Poisson–Boltzmann
RMS	root mean square
RMSD	root mean square deviation
SAS	solvent-accessible surface
UHBD	University of Houston Brownian Dynamics

References

1. B. K. Shoichet and I. D. Kuntz, *J. Mol. Biol.*, **221**, 327 (1991).
2. E. C. Meng, B. K. Shoichet, and I. D. Kuntz, *J. Comput. Chem.*, **13**, 505 (1992).
3. D. E. Koshland, Jr., *Proc. Natl. Acad. Sci. USA*, **44**, 98 (1958).
4. J. H. Arevalo, A. Hassig, E. A. Stura, M. J. Sims, M. J. Taussig, and I. A. Wilson, *J. Mol. Biol.*, **241**, 663 (1994).
5. M. Rarey, B. Kramer, T. Lengauer, and G. Klebe, *J. Mol. Biol.*, **261**, 470 (1996).
6. R. S. Judson, Y. T. Tan, E. Mori, C. Melius, E. P. Jaeger, A. M. Treasurywala, and A. Mathiowetz, *J. Comput. Chem.*, **16**, 1405 (1995).
7. K. P. Clark and Ajay, *J. Comput. Chem.*, **16**, 1210 (1995).
8. G. Jones, P. Willett, and R. C. Glen, *J. Mol. Biol.*, **245**, 43 (1995).
9. M. Zacharias, B. A. Luty, M. E. Davis, and J. A. McCammon, *J. Mol. Biol.*, **238**, 455 (1994).
10. A. R. Leach, *J. Mol. Biol.*, **235**, 345 (1994).
11. B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus, *J. Comput. Chem.*, **4**, 187 (1983).
12. J. Warwicker and H. C. Watson, *J. Mol. Biol.*, **157**, 671 (1982).
13. M. K. Gilson and B. H. Honig, *Proteins: Struct Funct. Genet.*, **4**, 7 (1988).
14. M. E. Davis and J. A. McCammon, *J. Comput. Chem.*, **10**, 386 (1989).
15. M. E. Davis and J. A. McCammon, *J. Comput. Chem.*, **11**, 401 (1990).
16. D. Bashford and M. Karplus, *Biochemistry*, **29**, 10219 (1990).
17. M. E. Davis and J. A. McCammon, *J. Comput. Chem.*, **12**, 909 (1991).
18. Z. Li and H. A. Scheraga, *Proc. Natl. Acad. Sci. USA*, **84**, 6611 (1987).
19. Z. Li and H. A. Scheraga, *J. Mol. Struct. (Theochem.)*, **179**, 333 (1988).
20. A. Caflisch, P. Niederer, and M. Anliker, *Proteins: Struct. Funct. Genet.*, **13**, 223 (1992).
21. A. Caflisch, S. Fischer, and M. Karplus, *J. Comput. Chem.*, **18**, 723 (1997).
22. J. H. Arevalo, E. A. Stura, M. J. Taussig, and I. A. Wilson, *J. Mol. Biol.*, **231**, 103 (1993).
23. D. W. Banner and P. Hadvary, *J. Biol. Chem.*, **266**, 20085 (1991).
24. F. H. Allen, S. Bellard, M. D. Brice, B. A. Cartwright, A. Doubleday, H. Higgs, T. Hummelink, B. G. Hummelink–Peters, O. Kennard, W. D. S. Motherwell, D. C. Rodgers, and J. R. Watson, *Acta Crystallogr.*, **B35**, 2331 (1979).
25. M. Zacharias, T. P. Straatsma, and J. A. McCammon. *J. Phys. Chem.*, **100**, 9025 (1994).
26. T. C. Beutler, A. E. Mark, R. C. van Schaik, P. R. Gerber, and W. F. van Gunsteren, *Chem. Phys. Lett.*, **222**, 529 (1994).
27. R. B. Hermann, *J. Phys. Chem.*, **76**, 2754 (1972).

28. D. Eisenberg and A. D. McLachlan, *Nature*, **319**, 199 (1986).
29. T. Ooi, M. Oobatake, M. Némethy, and H. A. Scheraga, *Proc. Natl. Acad. Sci. USA*, **84**, 3086 (1987).
30. J. Tomasi and M. Persico, *Chem. Rev.*, **94**, 2027 (1994).
31. W. C. Still, A. Tempczyk, R. C. Hawley, and T. Hendrickson, *J. Am. Chem. Soc.*, **112**, 6127 (1990).
32. D. Sitkoff, K. A. Sharp, and B. Honig, *J. Phys. Chem.*, **98**, 1978 (1994).
33. B. Lee and F. M. Richards, *J. Mol. Biol.*, **55**, 379 (1971).
34. C. Chothia, *Nature*, **248**, 338 (1974).
35. D. T. Edmonds, N. K. Rogers, and M. J. E. Sternberg, *Mol. Phys.*, **52**, 1487 (1984).
36. V. Mohan, M. E. Davis, J. A. McCammon, and B. M. Pettitt, *J. Phys. Chem.*, **96**, 6428 (1992).
37. B. A. Luty, M. E. Davis, and J. A. McCammon, *J. Comput. Chem.*, **13**, 768 (1992).
38. A. Caflisch, P. Niederer, and M. Anliker, *Proteins: Struct. Funct. Genet.*, **14**, 102 (1992).
39. N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, *J. Chem. Phys.*, **21**, 1087 (1953).
40. A. Miranker and M. Karplus, *Proteins: Struct. Funct. Genet.*, **11**, 29 (1991).
41. C. Tapparelli, R. Metternich, C. Ehrhardt, and N. S. Cook, *TIPS*, **14**, 366 (1993).
42. M. T. Stubbs and W. Bode, *Perspect. Drug Discov. Des.*, **1**, 431 (1993).
43. H. Brandstetter, D. Turk, H. W. Hoeffken, D. Grosse, J. Stuerzebecher, D. P. Martin, B. F. P. Edwards, and W. Bode, *J. Mol. Biol.*, **226**, 1085 (1992).
44. K. Hilpert, J. Ackermann, D. W. Banner, A. Gast, K. Gubernator, P. Hadvary, L. Labler, K. Müller, G. Schmid, T. Tschopp, and H. van de Waterbeemd, *J. Med. Chem.*, **37**, 3889 (1994).
45. D. R. Rich, In *Perspectives in Medicinal Chemistry*, B. Testa, E. Kyburz, W. Fuhrer, and R. Giger, eds., 1993, p. 13.
46. E. F. Meyer, I. Botos, L. Scapozza, and D. Zhang, *Perspect. Drug Discov. Des.*, **3**, 168 (1995).
47. C. M. Oshiro, I. D. Kuntz, and J. S. Dixon, *J. Comput.-Aided Mol. Des.*, **9**, 113 (1995).
48. H. J. Böhm and G. Klebe, *Angew. Chem. Int. Ed. Engl.*, **35**, 2588 (1996).
49. D. S. Goodsell and A. J. Olson. *J. Mol. Rec.*, **9**, 1 (1996).
50. P. Willett, *Trends Biotechnol.*, **13**, 516 (1995).
51. F. Jiang and S.-H. Kim, *J. Mol. Biol.*, **219**, 79 (1991).
52. R. M. Jackson and M. J. E. Sternberg, *J. Mol. Biol.*, **250**, 258 (1995).
53. H. J. Böhm, *J. Comput.-Aided Mol. Des.*, **6**, 61 (1992).
54. H. J. Böhm, *J. Comput.-Aided Mol. Des.*, **6**, 593 (1992).
55. A. Caflisch, A. Miranker, and M. Karplus, *J. Med. Chem.*, **36**, 2142 (1993).
56. H. J. Böhm, *J. Comput.-Aided Mol. Des.*, **8**, 243 (1994).
57. A. Caflisch, *J. Comput.-Aided Mol. Des.*, **10**, 372 (1996).
58. E. Skrzypczak−Jankun, V. E. Carperos, K. G. Ravichandran, A. Tulinsky, M. Westbrook, and J. M. Maraganore, *J. Mol. Biol.*, **221**, 1379 (1991).